REMARKS/ARGUMENTS

In view of the foregoing amendments and the following remarks, the applicants respectfully submit that the pending claims are not anticipated under 35 U.S.C. § 102 and are not rendered obvious under 35 U.S.C. § 103. Accordingly, it is believed that this application is in condition for allowance. **If, however, the Examiner believes that there are any unresolved issues, or believes that some or all of the claims are not in condition for allowance, the applicants respectfully request that the Examiner contact the undersigned to schedule a telephone Examiner Interview before any further actions on the merits.**

The applicants will now address each of the issues raised in the outstanding Office Action.

Objections

Claim 54 is objected to as depending from itself. Since claim 54 has been amended to depend from claim 53, this objection should be withdrawn.

Double Patenting Rejections

Claims 46 and 49 stand rejected on the ground of nonstatutory double patenting over claims 14-16 of U.S. Patent No. 6,658,423. Claim 47 stands rejected on the ground of nonstatutory obviousness-type double patenting over claim 14 of U.S. Patent No. 6,658,423. Claims 46 and 47 have been slightly amended to more clearly recite

the claimed invention. In any event, since a terminal disclaimer is filed herewith, this ground of rejection has been obviated and should be withdrawn.

Rejections under 35 U.S.C. § 102

Claim 48 stands rejected under 35 U.S.C. § 102 as being anticipated by U.S. Patent No. 6,360,215 ("the Judd patent"). The applicants respectfully request that the Examiner reconsider and withdraw this ground of rejection in view of the following.

Claim 48, as amended, is not anticipated by the Judd patent at least because the Judd patent does not teach ***a plurality of lists***, each of the plurality of lists containing ***elements*** of a document, nor does the Judd patent teach a data structure wherein a hash function is used to hash each of the elements to ***determine which one of the plurality of lists that each of the elements will be contained in***.

The Judd patent is directed to combating spamming (e.g., embedding unseen "decoy" words that are not relevant to the content of a document) of an indexing system or search engine system. (See, e.g., column 1, line 22 through column 2, line 3.) The Examiner cites various portions of columns 6 and 7 as teaching the invention of claim 48. The applicants respectfully disagree.

The cited portions of the Judd patent concern creating a word index which maps a word to a list of document identifiers (See, e.g., column 7, lines 3-5 and lines 57-64.) and a document index which maps a document identifier to a document URL and/or other characteristics

of a document such as title, summary, a hash of document contents, etc. (See, e.g., column 7, lines 5-9 and 41-57.) The Judd patent notes that words in the word index may be stored in hashed form for purposes of efficiency and speed. (See, e.g., column 7, line 65 through column 8, line 9.)

None of the foregoing teaches a data structure having (i) a first field for storing a document identifier, and (ii) ***a plurality of lists***, each of the plurality of lists containing ***elements*** of a document identified by the document identifier stored in the first field. That is, in the word index of the Judd patent, a word (which may be an element of a document) is mapped to a plurality of document identifiers. Thus, claim 48 is not anticipated by the Judd patent for at least this reason.

Further, the Judd patent does not teach such a data structure wherein a hash function is used to hash each of the elements ***to determine which one of the plurality of lists that each of the elements will be contained in.*** Instead, in the Judd patent, the word is simply hashed. The hash value is not used to determine which of a plurality of lists to include the word in. Thus, claim 48 is not anticipated by the Judd patent for at least this additional reason.

Claim 49 stands rejected under 35 U.S.C. § 102 as being anticipated by U.S. Patent No. 6,119,124 ("the Broder patent"). The applicants respectfully request that the Examiner reconsider and withdraw this ground of rejection in view of the following.

Claim 49 is not anticipated by the Broder patent because the Broder patent does not teach concluding that two documents are near-duplicates if ***any one*** fingerprint of one of the documents matches ***any one*** fingerprint of the other document, ***where each of the documents has at least two fingerprints***.

The Examiner contends that the Broder patent teaches this feature, citing column 10, lines 27-29. However, this section concerns reducing computational workload by eliminating (1) identical documents and (2) equivalent documents such that a cluster of documents does not include identical or equivalent documents. The Broder patent does so by (1) fingerprinting the entire document (for purposes of identifying identical documents) and (2) fingerprinting a canonical form of the document and/or a set of shingles of a document (for purposes of identifying equivalent documents) so that if two documents with identical fingerprints are encountered, only one is used in the clustering process. After clustering is completed, the eliminated documents are added back in. (See, e.g., column 10, lines 12-30.)

As can be appreciated from the foregoing, the fingerprints of entire documents (or of a canonical form of a document or of a set of shingles of a document) are not used to conclude whether or not two documents are ***near duplicates***. Rather, they are used in an optimization technique applied during clustering. (See, e.g., column 9, lines 59 and 60.) Furthermore, claim 49 recites that each of the documents includes ***at least two*** fingerprints. In the cited portion of the Broder patent, ***single*** fingerprints, representative of each document, are used to find identical (or lexically-equivalent or

shingle-equivalent) documents. Claim 49 has been amended to more clearly recite this feature. Thus, claim 49 is not anticipated by the Broder patent for at least the foregoing reasons.

Claims 50-54 stand rejected under 35 U.S.C. § 102 as being anticipated by U.S. Patent No. 6,873,982 ("the Bates patent"). The applicants respectfully request that the Examiner reconsider and withdraw this ground of rejection in view of the following.

Independent claims 50, 52 and 53 are not anticipated by the Bates patent because the Bates patent does not store a plurality of records, each of the records comprising (i) a first field for storing a document identifier, and (ii) a plurality of ***lists***, each of the plurality of lists containing ***elements*** of a document identified by the document identifier stored in the first field. Figure 3, element 355, as well as Figures 12a and 12b illustrate examples of the claimed data structure.

The Examiner contends that Figure 4 of the Bates patent teaches this feature. In particular, the Examiner contends that the keyword fields 106 teach the claimed "lists". Assuming, arguendo, that each of the keyword fields 106 of the Bates patent could be characterized as the claimed "lists", each of the keyword fields 106 does not contain ***elements*** of the document. Instead, each of the keyword fields 106 includes a ***single word*** of the document. Accordingly, independent claims 50, 52 and 53 are not anticipated by the Bates patent for at least this reason.

Rejections under 35 U.S.C. § 103

Claim 48 stands rejected under 35 U.S.C. § 103 as being obvious in view of the Bates and Judd patents. The applicants respectfully request that the Examiner reconsider and withdraw this ground of rejection in view of the following.

The Examiner alleges that Figure 4 of the Bates patent teaches a plurality of ***lists***, each of the plurality of lists containing ***elements*** of a document identified by the document identifier stored in the first field. However, as just discussed above with reference to claims 50, 52 and 53, the Bates patent does not teach this feature. Similarly, as discussed above with reference to the 102-based rejection of claim 48, the Judd patent does not teach this feature. Accordingly, claim 48 is not rendered obvious by the Bates and Judd patents for at least this reason.

Further, the Judd and Bates patents neither teach, not suggest, such a data structure wherein a hash function is used to hash each of the elements ***to determine which one of the plurality of lists that each of the elements will be contained in***. Thus, claim 48 is not rendered obvious by the Judd and Bates patents for at least this additional reason.

New claims

New claims 55, 58, 61 and 64 depend from claims 48, 50, 52 and 53, respectively, and further recite that each of the elements of a document is an element that has been extracted from the document. New claims 56, 59, 62, and

65 depend from claims 48, 50, 52 and 53, respectively, and further recite that each of the elements of a document is a predetermined one of (A) a predetermined number of words, (B) a predetermined number of sentences, (C) a predetermined number of characters, (D) a predetermined number of paragraphs, and (E) a predetermined number of sections. Finally, new claims 57, 60, 63, and 66 depend from claims 48, 50, 52 and 53, respectively, and further recite that each of the elements of a document partially overlaps another of the elements of the document. These new claims are supported, for example, by page 25, lines 11-18 of the specification. New claim 67 is similar to amended claim 48 and is allowable for the same reasons as discussed above with reference to claim 48.

Conclusion

In view of the foregoing amendments and remarks, the applicants respectfully submit that the pending claims are in condition for allowance. Accordingly, the applicants request that the Examiner pass this application to issue.

Respectfully submitted,

July 24, 2006

John C. Pokotylo, Attorney
Reg. No. 36,242
Tel.: (732) 542-9070

CERTIFICATE OF MAILING under 37 C.F.R. 1.8(a)

I hereby certify that this correspondence is being deposited on **July 24, 2006** with the United States Postal Service as first class mail, with sufficient postage, in an envelope addressed to Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

John C. Pokotylo

36,242
Reg. No.